



HAL
open science

Mean field Limit of Non-Smooth Systems and Differential Inclusions

Nicolas Gast, Bruno Gaujal

► **To cite this version:**

Nicolas Gast, Bruno Gaujal. Mean field Limit of Non-Smooth Systems and Differential Inclusions. [Research Report] RR-7315, 2010. inria-00491859v2

HAL Id: inria-00491859

<https://inria.hal.science/inria-00491859v2>

Submitted on 22 Dec 2010 (v2), last revised 12 Mar 2012 (v4)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Mean field Limit of Non-Smooth Systems and Differential Inclusions

Nicolas Gast — Bruno Gaujal

N° 7315 — version 2

initial version June 2010 — revised version September 2010

Thème NUM



R
apport
de recherche

Mean field Limit of Non-Smooth Systems and Differential Inclusions

Nicolas Gast, Bruno Gaujal

Thème NUM — Systèmes numériques
Équipe-Projet MESCAL

Rapport de recherche n° 7315 — version 2 — initial version June 2010 — revised version
September 2010 — 21 pages

Abstract: In this paper, we study deterministic limits of Markov processes made of several interacting objects. While most classical results assume that the limiting dynamics has Lipschitz properties, we show that these conditions are not necessary to prove convergence to a deterministic system.

We show that under mild assumptions, the stochastic system converges to the set of solutions of a differential inclusion and we provide simple way to compute the limiting inclusion. When this differential inclusion satisfies a one-sided Lipschitz condition, there exists a unique solution of this differential inclusion and we show convergence in probability with explicit bounds.

This extends the applicability of mean field techniques to systems exhibiting threshold dynamics such as queuing systems with boundary conditions or controlled dynamics. This is illustrated by applying our results to several types of systems: fluid limits of priority queues, best response dynamics in games, push-pull queues with a large number of sources and a large number of servers and self-adapting computing systems.

Key-words: Mean Field, Differential inclusions, Queuing systems

Limite en Champ Moyen de systèmes non-réguliers et inclusions différentielles

Résumé : Dans cet article nous étudions le comportement limite de processus de Markov composés de multiples objets en interaction. Alors que la plupart des résultats classiques supposent que la dynamique limite a des propriétés de régularité de type Lipschitz, nous montrons que ces conditions ne sont pas nécessaires pour prouver la convergence vers un système déterministe. Nous montrons que sous des hypothèses assez faibles, le système stochastique converge vers la solution d'une inclusion différentielle et nous donnons un moyen simple d'obtenir cette inclusion différentielle limite. Quand elle satisfait une condition de semi-Lipschitz, il existe une solution unique à cette inclusion différentielle et nous montrons la convergence en probabilité en donnant des bornes explicites.

Cela permet d'étendre l'applicabilité des techniques de champs moyen à des systèmes avec des dynamiques à seuil, comme c'est le cas pour des systèmes de files d'attente. Ceci est illustré par l'application de nos résultats à des files "push-pull" avec un grand nombre de sources de paquets et un grand nombre de serveurs, qui constituent des modèles naturels de systèmes de calculs à volontaires.

Mots-clés : Champ Moyen, Inclusions différentielles, Systèmes de files d'attente

1 Introduction

Let us consider a discrete stochastic dynamical system composed by a large number of objects. Under classical smoothness assumptions, there exist general results that show that the limiting system (when the number of objects goes to infinity) can be described by a system of deterministic ordinary differential equations

$$\dot{y}(t) = f(y(t)). \quad (1)$$

See for example [4] and the references therein for examples of such convergence results, known as mean field limits. In most cases, the limiting drift function f in (1) is assumed to be Lipschitz. This strong condition hampers the applicability of these results in many practical cases, in particular, for systems exhibiting thresholds dynamics or with boundary conditions. This paper studies the limiting behavior of such a system when the dynamics is non-smooth. Let us consider a simple queuing system with one buffer and many processors that can serve one packet per unit of time, on average. If y denotes the number of packets in the queue, then the average decrease of y is one packet per unit of time (under a proper rescaling of time) if the queue is non-empty (*i.e.* $y > 0$) and zero if the queue is empty. This leads to a deterministic limit behavior:

$$\dot{y}(t) = -1 \text{ if } y(t) > 0 \quad \text{and} \quad \dot{y}(t) = 0 \text{ if } y(t) = 0. \quad (2)$$

This dynamics is not continuous and therefore non Lipschitz which makes the classical approach inapplicable in that case.

Actually, most work using mean field limits for networks do not involve queues or when they do, the number of queues scale with the number of objects (as in [7]), or convergence is obtained using *ad hoc* proofs (see for example [2]).

In the case of a non-continuous right-hand side, the differential equation (1) is not well-defined since there exist no function y that is differentiable and that satisfies (2). The proper way to define solutions of (2) is to use differential inclusions (DI) instead. Equation (1) is replaced by the following equation

$$\dot{y}(t) \in F(y(t)). \quad (3)$$

where F is a set-valued mapping: if $y \neq 0$ then $F(y) = \{-1\}$ and $F(0) = \{u : -1 \leq u \leq 0\}$. Of course a differential inclusion problem may (or may not) have multiple solutions.

In the following, we will provide generic convergence results of mean field type that show that under few conditions on the system, its behavior converges to the solutions of a differential inclusion (3) (Theorem 5). This result is generic and does not require any Lipschitz property on the function F . In particular, it shows that when (3) has a unique solution, the behavior of the system converges to it. Moreover, we also show that when F satisfies a one-sided Lipschitz condition (9), we can bound the gap with the limiting dynamics with explicit bounds (Theorem 7).

The rest of the paper is organized as follows. In Section 2, we briefly describe the general framework of our work and we give several examples. We will briefly recall some definitions and properties of differential inclusions in Section 2.1. Section 3 states the main theoretical results, Section 4 extends the results to time continuous Markov chains and Section 5 describes several application examples.

2 Description of the Model and Notations

We consider of system composed by N objects evolving in a finite state space $\mathcal{S} = \{1 \dots S\}$. Time is discrete and the state of object n at time step k is denoted $X_n^N(k)$ and $\mathcal{X}^N(k) \stackrel{\text{def}}{=} (X_1^N(k), \dots, X_N^N(k))$. The objects all evolve in a common environment, called the *context*. The context at time step k is denoted by $C^N(k) \in \mathbb{R}^d$. The state of the global system at time k is $(\mathcal{X}^N(k), C^N(k))$. We denote by $M^N(k)$ the empirical measure associated with the N objects:

$$M^N(k) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{n=1}^N \delta_{X_n^N(k)}.$$

Since an object has S possible states, $M^N(k)$ can be represented by a vector with S components, its i th component being the proportion of objects in state i :

$$M_i^N(k) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{n=1}^N \mathbf{1}_{X_n^N(k)=i}.$$

The system $(M^N(k), C^N(k))_k$ is assumed to be a Markov chain. In particular, this is true if the objects all have a Markovian dynamics, if the evolution of the context is deterministic and if the law of the whole system is invariant by any permutation of the N objects. The state space of this Markov chain is included in \mathbb{R}^{S+d} . To simplify the notations, we call $Y^N(k) \stackrel{\text{def}}{=} (M^N(k), C^N(k))$.

If at time k , the system is in state $Y^N(k) = y$, then the expected difference between $Y^N(k+1)$ and $Y^N(k)$ is called the *drift* and is denoted $f^N(y)$:

$$f^N(y) \stackrel{\text{def}}{=} \mathbb{E}(Y^N(k+1) - Y^N(k) | Y^N(k) = y). \quad (4)$$

This only defines f^N on the state space of Y^N , which is a subset of \mathbb{R}^{d+S} . Outside the state space of Y^N , $f^N(y)$ is not defined at this point.

We assume that as N grows, the drift vanishes with speed $I(N)$. This means that there exists a function $I(N)$, called the *intensity* of the model, a set-valued function F and a continuation of the function f^N on all \mathbb{R}^{d+S} such that

- The intensity vanishes: $\lim_{N \rightarrow \infty} I(N) = 0$.
- f^N converges uniformly to F in the following sense: there exists $J(N)$ with $\lim_{N \rightarrow \infty} J(N) = 0$ and a function f such that $f(y) \in F(y)$ for all $y \in \mathbb{R}^{d+S}$, satisfying

$$\sup_{y \in \mathbb{R}^{d+S}} \left\| \frac{f^N(y)}{I(N)} - f(y) \right\| \leq J(N). \quad (5)$$

where $\|\cdot\|$ denotes the L^2 norm.

Actually, the following results (Theorems 5 and 7) remain valid under a weaker convergence condition of f^N to F . Instead of assuming that f^N has a limit f , one could simply assume that

$$\sup_{y \in \mathbb{R}^{d+S}} \inf_{u \in F(y)} \left\| \frac{f^N(y)}{I(N)} - u \right\| \leq J(N). \quad (6)$$

All the foregoing proofs would hold by replacing f by any h achieving the infimum in the above equation (6). However, in most practical cases (and in all the examples of this paper), the rescaled drift $f^N/I(N)$ converges indeed to a (non-continuous) function f , so that we focus on that case in the following.

Finally, we assume that the second order of the drift is bounded. This means that there exists $b > 0$ such that:

$$\mathbb{E} \left(\left\| \frac{Y^N(k+1) - Y^N(k) - f^N(Y^N(k))}{I(N)} \right\|^2 \right) \leq b. \quad (7)$$

All these assumptions are more or less necessary to use a mean field approach. The notable important feature of our model is that we do not assume any regularity property on the function f .

2.1 Differential inclusions

In Section 3, we will see that under mild conditions, the system described by $(M^N(\cdot), C^N(\cdot))$ converges to the solutions of a deterministic differential inclusion. In this section, we recall the main concepts on differential inclusions. For a more complete description, the reader is referred to [1]. In all that follows, $\langle x, y \rangle$ denotes the classical inner-product on \mathbb{R}^{d+S} and $\|x\| = \sqrt{\langle x, x \rangle}$ (L^2 norm) and $\|A\| = \sup_{x \in A} \|x\|$.

Definition 1. Consider a differential inclusion problem:

$$\dot{y}(t) \in F(y(t)), \quad y(0) = y_0, \quad (8)$$

where F is a set-valued function mapping each point $y \in \mathbb{R}^{d+S}$ to a set $F(y) \subset \mathbb{R}^{d+S}$. Let $I \subset \mathbb{R}$ be an interval with $0 \in I$. A function $y : I \rightarrow \mathbb{R}^{d+S}$ is a solution of the DI (8) with initial condition $y(0) = y_0$ if there exists a function $\varphi : I \rightarrow \mathbb{R}^{d+S}$ such that:

- (i) for all $t \in I$: $y(t) = y_0 + \int_0^t \varphi(s) ds$;
- (ii) for almost every (a.e.) $t \in I$: $\varphi(t) \in F(y(t))$.

In particular, (i) is equivalent to say that y is absolutely continuous. (i) and (ii) imply that y is differentiable at almost every $t \in I$ with $\dot{y}(t) \in F(y(t))$.

Definition 2 (Upper Semi-Continuous (USC)). The function F is upper semi-continuous (USC) if for any $y \in \mathbb{R}^{d+S}$, $F(y)$ is a non-empty closed, convex and bounded set and if for any open set O containing $F(y)$, there exists a neighborhood V of y such that $F(V) \subset O$.

Definition 3 (One-Sided Lipschitz (OSL)). A set-valued function F is one-sided Lipschitz (OSL) with constant L if for all $y, \bar{y} \in \mathbb{R}^{d+S}$ and for all $u \in F(y)$ $\bar{u} \in F(\bar{y})$:

$$\langle y - \bar{y}, u - \bar{u} \rangle \leq L \|y - \bar{y}\|^2. \quad (9)$$

These two definitions give sufficient conditions for the existence (resp. uniqueness) of solutions for the differential inclusion (8). We recall the following results.

Proposition 4 (Theorems 2.2.1 and 2.2.2 of [8]).

- If F is USC and if there exists c such that $\|F(x)\| \leq c(1 + \|x\|)$ then for all initial condition y_0 , (8) has at least one solution on $[0; \infty)$ with $y(0) = y_0$.
- If F is OSL, then for all $T > 0$, there exists at most one solution of (8) on $[0; T]$.

Of course (USC) and (OSL) combined insure that the DI has a unique solution.

2.2 A set valued function for the drift

In our framework, if $f^N(\cdot)$ is the drift of the system defined in Equation (4), we define an adapted set-valued function F by:

$$F(y) = \bigcap_{\epsilon > 0} \lim_{K \rightarrow \infty} \bigcup_{N \geq K} \overline{\text{conv}} \left\{ \frac{f^N(z)}{I(N)} : \|z - y\| \leq \epsilon \right\}. \quad (10)$$

where $\overline{\text{conv}}(A)$ is the closure of the convex hull of a set A . It should be clear that Equation (10) defines a set-valued function F such that for all y : $\lim_{N \rightarrow \infty} \inf_{u \in F(y)} \|f^N(y)/I(N) - u\| = 0$. Therefore, Equation (6) is satisfied as soon as this convergence holds uniformly in y . Furthermore, the function F defined in (10) is USC.

If we assume instead that the drift f^N converges uniformly to some function f , meaning that $\|f^N(y)/I(N) - f(y)\| \leq J(N)$, the set valued function F associated to the drift can be defined as:

$$F(y) = \bigcap_{\epsilon > 0} \overline{\text{conv}} \{f(z) : \|z - y\| \leq \epsilon\}. \quad (11)$$

In that case, if the function f is continuous in a point y , then F is a singleton: $F(y) = \{f(y)\}$ while if f is discontinuous in y , $F(y)$ is the convex closure of the limit set of f at point y .

3 Convergence results

This section contains the two main theoretical contributions of this paper. The first one is Theorem 5 that shows that if F is USC then the stochastic system converges to the set of solutions of the differential inclusion. In particular, this implies that if the differential inclusion has only one solution, the stochastic system converges to this solution (Corollary 6).

This theorem does not give any bound on the speed of convergence. In fact, without further conditions, the convergence may be arbitrarily slow. However, when the differential inclusion is both USC and OSL, the the solution of the differential equation is unique and the speed of convergence can be lower-bounded (Theorem 7).

The rest of this section is organized as follows. We first state the two convergence results and then discuss their applicability. the proofs of the two theorems are postponed in the appendix. The proof of the first theorem is similar to classical proofs for the existence of solutions of differential inclusions while the second one is more similar to classical proofs of the convergence of stochastic approximation algorithms like in [4, 7].

Let us now recall that at time step k , the system is in state $Y^N(k) \stackrel{\text{def}}{=} (M^N(k), C^N(k))$ and let us denote $\bar{Y}^N(t)$ the piecewise affine interpolation of $Y^N(k)$ when we scale time by a factor $I(N)$. Let us first define the continuous function $\tilde{Y}^N(t)$ as the component-wise linear interpolation of $\{Y^N(k)\}_{k \in \mathbb{N}}$: for all i ,

$$\tilde{Y}_i^N(t) \stackrel{\text{def}}{=} (1 - \alpha)Y_i^N(\lfloor t \rfloor) + \alpha Y_i^N(\lceil t \rceil),$$

with $\alpha = t - \lfloor t \rfloor$.

Then, by scaling time, $\bar{Y}^N(t) \stackrel{\text{def}}{=} \tilde{Y}^N(t/I(N))$.

Let us denote by $\mathcal{S}_T(y_0)$ the set of the solutions of the DI (8) starting from $y(0) = y_0$, we can show that $\bar{Y}(\cdot)$ converges (in probability) to $\mathcal{S}_T(y_0)$. More precisely, the following theorem holds.

Theorem 5. *Let F be the limit of the drift defined by Equation (5). Assume that the drift satisfies (7) and that*

- F is USC and there exists $c > 0$ s.t. $\|F(y)\| \leq c(1 + \|y\|)$,
- $Y^N(0) \xrightarrow{\mathcal{P}} y_0$ (convergence in probability).

Then for all $T > 0$:

$$\inf_{y \in \mathcal{S}_T(y_0)} \sup_{0 \leq t \leq T} \|\bar{Y}^N(t) - y(t)\| \xrightarrow{\mathcal{P}} 0.$$

Proof. The proof is given in A.1. □

This theorem shows that if N is large enough, the trajectory of the stochastic system \bar{Y}^N on $T/I(N)$ steps is close to a solution of the differential inclusion (8) over $[0, T]$. In general a differential inclusion may have multiple solutions. Here, \bar{Y}^N may converge to any solution of the DI, depending on its random innovations, making this result rather inefficient for performance evaluation. This result is of greater interest if the DI starting from y_0 has a unique solution: $\mathcal{S}_T(y_0) = \{y\}$. In that case, as a direct corollary of the preceding result, \bar{Y}^N converges in probability to y on all intervals $[0; T]$.

Corollary 6. *Under the conditions of Theorem 5, if the DI (8) has a unique solution y , then for all T :*

$$\sup_{0 \leq t \leq T} \|\bar{Y}^N(t) - y(t)\| \xrightarrow{\mathcal{P}} 0.$$

In some cases, like the example of push-pull queues described in Section 5, the limiting differential inclusion clearly has a unique solution which makes the preceding corollary directly applicable. The main drawback of the previous theorem is that it does not give any insight on the speed of convergence on the stochastic system towards its limit.

This limitation can be overcome when the function F satisfies the one-sided Lipschitz condition (9). Firstly, this ensures the uniqueness of the solution. Secondly, one can get precise bounds on the gap between the stochastic system and its limit in that case.

Theorem 7. Under the conditions of Theorem 5 and if F is OSL with constant L , then the DI (8) has a unique solution y and there exist constants A_T, B_T, C_T depending only on T, L and c such that for all ϵ ,

$$\mathcal{P} \left(\sup_{0 \leq t \leq T} \|Y^N(t) - y(t)\| \geq \|Y^N(0) - y(0)\| e^{LT} + \sqrt{I(N)}A_T + \sqrt{J(N)}B_T + \epsilon \right) \leq \frac{I(N)}{\epsilon^2} C_T.$$

The constants A_T, B_T and C_T are given by

$$\begin{aligned} A_T &\stackrel{\text{def}}{=} e^{2LT} \sqrt{\frac{I(N)}{6} c^2 (1 + K_T) + \frac{K_T}{2L} (c(1 + K_T) + J(N))}; \\ B_T &\stackrel{\text{def}}{=} \frac{\sqrt{K_T} e^{2LT}}{\sqrt{L}}; \\ C_T &\stackrel{\text{def}}{=} \frac{e^{2LT}}{2L} (4c^2 (1 + K_T)^2 + b) \end{aligned}$$

with $K_T = \max\{\|y(0)\| + c, \|y_0^N\| + (c + J(N))T, \|Y_0^N\| + (c + J(N))T + \sqrt{bT}\} e^{ct}/c$ and c is defined in Theorem 5 and b in Equation (7).

Proof. The proof is given in A.2. □

Note that if $F(\cdot)$ is bounded by some K_F the terms in $c(1 + K_T)$ can be replaced by K_F . This is in particular true if Y^N is constrained to stay in a compact space of \mathbb{R}^{d+S} or if the drift is bounded for all $y \in \mathbb{R}^{d+S}$.

These constants are of a similar order than bounds that can be obtained in the case where f is Lipschitz (see [7]). However, the convergence speed with respect to N is in $O(\sqrt{I(N)})$ (compared with $O(I(N))$ is the Lipschitz case).

4 Extension to Non-smooth Density Dependent Population Processes

In this section, we show that our results can be adapted to the case of continuous time Markov chains and in particular for the famed model of density dependent population processes of Kurtz [9]. The two theorems 5 and 7 can be transposed in this case.

Let D^N be a continuous Markov chain on $\frac{1}{N}\mathbb{Z}^d$ ($d \geq 1$) for $N \geq 1$. D^N is called a *density dependent population process* if there exists a set $\mathcal{L} \subset \mathbb{Z}^d$ (with $0 \notin \mathcal{L}$), such that for each $\ell \in \mathcal{L}$ and $x \in N^{-1}\mathbb{Z}^d$, the rate of transition from x to $x + \ell/N$ is $N\beta_\ell(x) \geq 0$, where $\beta_\ell(\cdot)$ does not depend on N . The i th component of $D^N(t)$, $D_i^N(t)$ can be seen as the density of individuals of a population that are in state i , hence the name, and a transition changes the number of individuals in state i by the quantity ℓ_i .

Let us assume that $\sum_{\ell \in \mathcal{L}} \sup_{x \in \mathbb{Z}^d} \|\beta_\ell(x)\| = \tau < \infty$ and let us define $f(x) = \sum_{\ell \in \mathcal{L}} \beta_\ell(x)\ell$ (in the following, we assume that this sum is well-defined). If f is Lipschitz, it is well-known that if $\lim_{N \rightarrow \infty} D^N(0) = d(0)$ in probability, then there exists a differentiable function $d(\cdot)$ such that $\lim_{N \rightarrow \infty} \sup_{t \leq T} \|D^N(t) - d(t)\| = 0$ in probability (see [9]).

Using uniformization of the Markov chain and the results from Section 3 we now show that this convergence still holds for general drifts, replacing f by its set-valued counterpart F , defined in (11).

Theorem 8. If $\sup_{x \in \mathbb{Z}^d} \sum_{\ell \in \mathcal{L}} \beta_\ell(x) = \tau < \infty$, if $\sum_{\ell \in \mathcal{L}} \|\ell\|^2 \beta_\ell^2(y) < b < \infty$ and if F is USC and satisfies $\|F(x)\| \leq c(1 + \|x\|)$, then for all $T > 0$:

$$\inf_{d \in \mathcal{S}_T(y_0)} \sup_{0 \leq t \leq T} \|D^N(t) - d(t)\| \xrightarrow{\mathcal{P}} 0.$$

where $\mathcal{S}_T(y_0)$ is the solution set of the DI (8) starting in y_0 .

Moreover, if F is OSL of constant L , then

$$\mathcal{P} \left(\sup_{0 \leq t \leq T} \|D^N(t) - d(t)\| \geq \|D^N(0) - d(0)\| e^{LT} + \frac{1}{\sqrt{N}} A_T + (1 + c(1 + K_T^y))\epsilon \right) \leq \frac{C_T + \tau}{N\epsilon^2}.$$

where A_T, C_T and K_T are defined as in Theorem 7 with $I(N) = N^{-1}$, and $d(\cdot)$ is the unique solution of the differential inclusion (8).

Proof. Since $\tau < \infty$, the rate of transition of $D^N(\cdot)$ is bounded by $N\tau$. Using uniformization of continuous time Markov chain (see [12] for example), there exists a Poisson process Λ^N of rate $N\tau$ and a discrete time Markov chain $Y^N(\cdot)$ such that $D^N(t) = Y^N(\Lambda^N(t))$ and Y^N and Λ^N are independent. Moreover, for all x and $\ell \in \mathcal{L}$,

$$\begin{aligned} \mathcal{P} \left(Y^N(k+1) = x + \frac{\ell}{N} \mid Y^N(k) = x \right) &= \frac{1}{\tau} \beta_\ell(x), \\ \mathcal{P} (Y^N(k+1) = x \mid Y^N(k) = x) &= 1 - \frac{1}{\tau} \sum_{\ell \in \mathcal{L}} \beta_\ell(x). \end{aligned}$$

For all $y \in \mathbb{R}^d$, the drift of $Y^N(\cdot)$ is $\mathbb{E}(Y^N(k+1) - Y^N(k) \mid Y^N(k) = y) = (N\tau)^{-1} f(y)$. Moreover, $\mathbb{E} \left(\|Y^N(k+1) - Y^N(k) - f(y)\|^2 \mid Y^N(k) = y \right) \leq (N\tau)^{-2} \sum_{\ell \in \mathcal{L}} \|\ell\|^2 \beta_\ell(y) < b$. Therefore, $Y^N(k)$ satisfies the conditions of Theorem 5. Moreover, F also satisfies the conditions of Theorem 5 which shows that $\inf_{y \in \mathcal{S}_T(y_0)} \sup_{t \leq T} \|Y^N(tN) - y(t)\| = 0$. When F satisfies the OSL condition of Theorem 7, we further get the result with explicit bounds.

As Λ^N is a Poisson process of rate $N\tau$, $|\Lambda^N(t) - tN\tau|^2$ is a sub-martingale and by Doob's inequality, $\mathcal{P}(\sup_{t \leq T} |\Lambda^N(t) - tN\tau| \geq N\tau\epsilon) \leq \mathbb{E} \left(|\Lambda^N(T) - TN\tau|^2 \right) / (N\tau\epsilon)^2 = (TN\tau) / (N\tau\epsilon)^2 = T / (N\tau\epsilon^2)$. If y is a solution of the DI (8) on $[0; T]$, for all $t, s \in [0, T]$, $\|y(t) - y(s)\| \leq c(1 + K_T^y) |t - s|$ where K_T^y is defined in Lemma 11. This shows that if y is a solution of the differential inclusion, with probability greater than $T / (N\tau\epsilon^2)$, we have:

$$\begin{aligned} \|D^N(t) - y(t)\| &= \|Y^N(\Lambda^N(t)) - y(t)\| \\ &\leq \left\| Y^N(\Lambda^N(t)) - y \left(\frac{\Lambda^N(t)}{N\tau} \right) \right\| + \left\| y \left(\frac{\Lambda^N(t)}{N\tau} \right) - y(t) \right\| \\ &\leq \left\| Y^N(\Lambda^N(t)) - y \left(\frac{\Lambda^N(t)}{N\tau} \right) \right\| + c(1 + K_T^y)\epsilon. \end{aligned}$$

which concludes both parts of the theorem. \square

5 Examples

5.1 Fluid limit of a system of parallel queues

Fluid limits are very popular for studying the dynamics of flows in a queuing network [6]. In this section, we show that our framework can be applied directly to prove the convergence of a queuing network to its corresponding fluid limit. Moreover, the differential inclusion approach allows us to compute the limiting dynamics as a direct application of Theorem 8.

We consider a simple model of a system composed by one server. There are two classes of customers: customers with high priority (HP) and customers with low priority (LP). High priority customers and low priority customer enter the system with rate λ (independently). If there are one or more HP customers in the queue, the server serves HP customers at rate 3λ . Only when there is no HP customer in the HP queue, can the server serve LP customers (with rate 3λ). Let $C_1(t)$ and $C_2(t)$ are the numbers of high priority and low priority customers at time t . This model

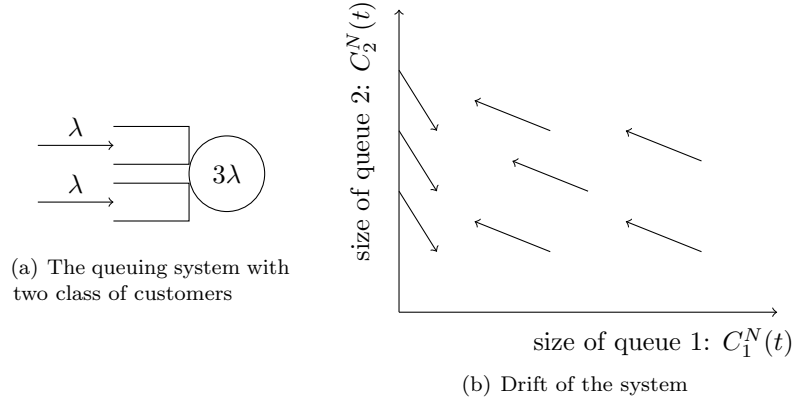


Figure 1: System and the corresponding drift

is depicted in Figure 1(a) where each queue corresponds to a class of customers. The fluid limit of this system corresponds to the limiting behavior of the system when the initial state is scaled by a factor going to infinity.

This model fits in our density dependent population process framework by considering a system with 0 object and only a shared resource $C^N = (C_1^N, C_2^N)$. At time 0, let us define an arbitrary initial state with, say, $4N$ high priority customers and N low priority customers. Thus $C^N(0) \stackrel{\text{def}}{=} (4, 1)$. It should be clear that the expected variation of $C^N(t)$ is $f^N(c_1, c_2) = \frac{1}{N}(-2\lambda, \lambda)$ if $C_1^N(t) > 0$ and $f^N(c_1, c_2) = \frac{1}{N}(\lambda, -2\lambda)$ if $C_1^N(t) = 0$. Therefore, the intensity of the system is $I(N) = \frac{1}{N}$ and the limit of the drift is $f(c_1, c_2) = Nf^N(c_1, c_2)$ and is depicted in Figure 1(b).

As shown on Figure 1(b), the drift is constant for all $C_1^N > 0$ but is discontinuous for $C_1^N = 0$. Because of this discontinuity, there is no absolutely continuous function c such that $\dot{c}(t) = f(c)$ almost everywhere: the axis $c_1 = 0$ both attracts the trajectories from $c_1 > 0$ and repulses the trajectories starting from $c_1 = 0$.

However, the intuition tells us that when $C_1^N(t) = 0$, $C_1^N(t)$ should remain close to 0 and that the server should serve in average λ HP customers and 2λ LP customers. To show that this intuition is correct, let us compute the set-valued function F corresponding to f (defined as in Equation (10)). For $c_1 > 0$, $F(c)$ is single-valued and $F(c) = \{f(c)\}$. For $c_1 = 0$, $F(c)$ is the convex hull of the vectors $\{(p, -2p), (-2p, p)\}$ and corresponds to the dashed line of Figure 2(a).

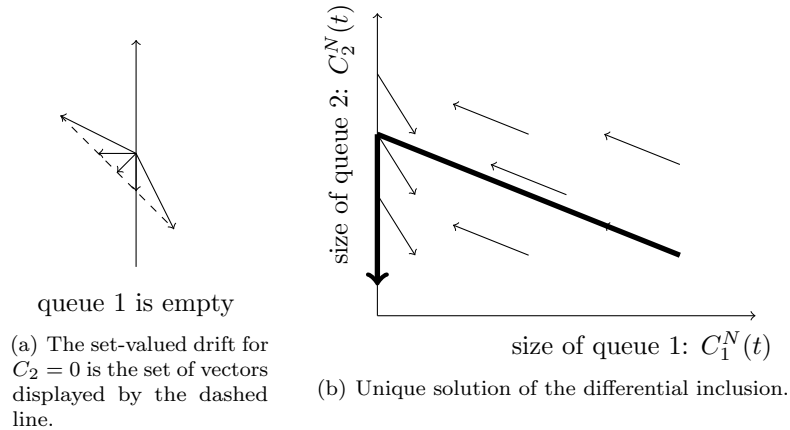


Figure 2: Convex hull of the drift at $C_2 = 0$ and unique solution of the fluid limit.

It should be clear that the model satisfies all the hypothesis of Theorem 5 and that the differential inclusion has a unique solution, depicted in Figure 2(b). Although this result can be shown directly, our framework provide an easy way to construct the fluid limit and prove the convergence of the original process.

5.2 Best Response in Rock-Paper-Scissors

The second example is taken from game theory and shows that the best response dynamics typically leads to a non-smooth limit system. In this case the limiting system will be USC and will have a unique solution but it will not satisfy a one-sided Lipschitz condition.

We consider a set of N individuals playing the game of rock-paper-scissors. The state of one individual can either be rock, paper or scissors. At each time step, 2 players are chosen at random. The first player cannot change its choice while the second can decide to play one of the three choices. The goal of the second player is to beat player one (the rule of the game is rock beats scissors, scissors beat paper and paper beats rock).

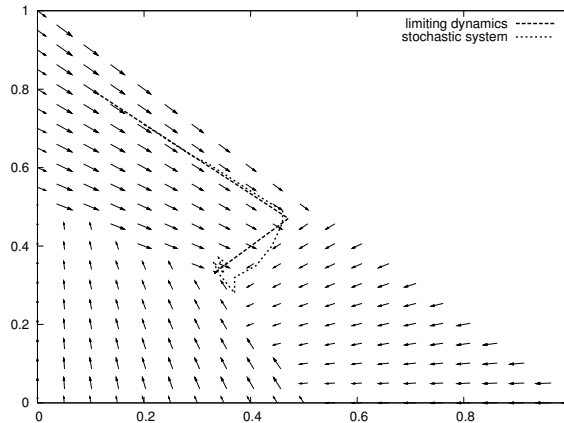


Figure 3: (Vector field for the best response dynamics. The first (second) coordinate M_1 (M_2) is the proportion of Rocks (Paper). The proportion of Scissors is $M_3 = 1 - M_1 - M_2$. Two trajectories starting in $M(0) = (.8, .1, .1)$ are shown. One corresponding to a stochastic system with $N = 10$ players and one for the deterministic limit.

The best response of player two is the following. If there is a majority of rock (resp. paper or scissors), the second player should play paper (resp. scissor or rock). The corresponding limiting dynamics is drawn on Figure 3. The drift is not Lipschitz on the frontiers and the limit dynamics is non-smooth. The differential inclusion is USC but not OSL. However, for all initial conditions, the solution of the differential inclusion is unique and has at most one cusp point before converging in finite time to its unique attractor $(1/3, 1/3, 1/3)$, at constant speed.

5.3 Volunteer Computing

Here, we consider a model of a volunteer computing system, such as BOINC <http://boinc.berkeley.edu/>. The system is made of a single buffer and N desktop machines, offered by their owners (volunteers), that serve the packets of this buffer. However, as soon the owner of a processor wants to use it, she preempts it and the processor becomes unavailable for the computing system. As for the incoming packets, they are assumed to arrive in the buffer according to a Poisson process at rate λ . These kinds of systems are often called push/pull models: The distributed applications *push* jobs to a central server that stores them in a buffer and whenever a processor becomes available, it *pulls* a job from the buffer and executes it.

Such systems fit our density dependent population process framework. The context $C(t)$ represents the size of the buffer while the N objects represent both the applications sending jobs

and the hosts executing them. The state of a host is its availability and its idleness (whether it is executing a job or not). The non-smooth part of the dynamics comes from the buffer size. When $C(t) > 0$, if a host asks for a job, it gets it with probability one while when $C(t) = 0$, a host asking for a job will get nothing. In that case, one can show that this dynamics satisfies the conditions of Theorem 8 that can be used to study the limiting behavior of the system when the number of hosts and applications grows.

In the simplest case, the intensity of the system is $I(N) = 1/N$ and an application sends a job to the system at rate λ while jobs are completed at rate μ by each server. To represent the communication delays, every host gets jobs at rate γ . It becomes unavailable with rate p_u , and available with rate p_a if $C(t) > 0$ and 0 otherwise. If b, a, u denote respectively the proportion of busy, available and unavailable hosts, the limiting system is described by a DI:

$$\begin{aligned} \dot{b}(t) &= -\mu b(t) + \gamma a(t) \mathbf{1}_{C(t) > 0} \\ \dot{a}(t) &= \mu(t) b(t) + p_a u(t) - \gamma a(t) \mathbf{1}_{C(t) > 0} \\ \dot{u}(t) &= -p_a u(t) + p_u a(t) \\ \dot{C}(t) &= -\gamma a(t) \mathbf{1}_{C(t) > 0} + \lambda \mathbf{1}_{C(t) < C_{\max}}. \end{aligned}$$

The formal DI is obtained by replacing $a(t) \mathbf{1}_{C(t) > 0}$ by the singleton $\{\gamma a(t)\}$ if $C(t) > 0$ and the interval $[0; \gamma a(t)]$ when $C(t) = 0$.

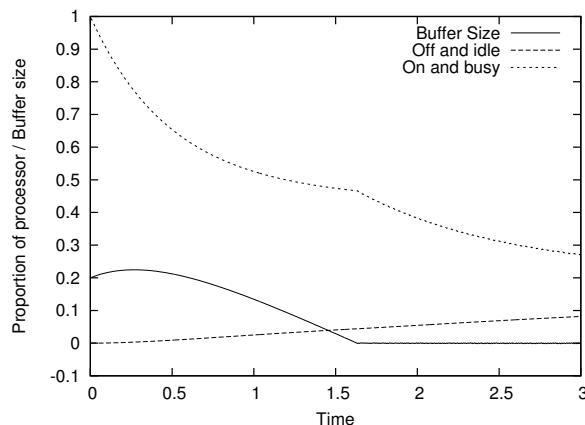


Figure 4: Limit dynamics of a volunteer computing system. A non-differentiable point occurs when the buffer becomes empty.

At time $t = 0$, we consider that the size of the buffer is $C(0) = .2$ and that all processors are available and are serving a job. The behavior of the system is represented in Figure 4. One can see that there is a point of non-differentiability in the behavior of the system when the size of the buffer reaches 0.

5.4 Volunteer Computing: Day and Night Scenario

We now consider a similar model as the previous one except that the processors follow a day and night behavior. We consider that some of the processors are turned off at night. Therefore, the availability is larger during the day (between 7am and 5pm) than at night.

The limiting dynamics can be represented by a differential inclusion that depends (not continuously) on time:

$$\frac{\partial y(t)}{\partial t} \in F(y(t), t). \quad (12)$$

In the previous theorems, we assumed that the function F was time-invariant. There are three ways to tackle this problem. The first one is to adapt the proofs to the time dependent case. An other idea that can be used here is the fact that for all finite interval of time, there is only a finite

number of discontinuity points (7am, 5pm, ...), and to apply the convergence results on a first sub-interval $[0; 7am]$. Using the fact that $Y^N(7) \rightarrow y(7)$, the convergence holds on $[7am, 5pm]$, and so forth. Yet another solution is to write $Z(t) \stackrel{\text{def}}{=} (Y(t), t)$. Then the differential inclusion (12) can be written:

$$\frac{\partial z(t)}{\partial t} \in (F(z(t)), 1). \quad (13)$$

The fact that F is not continuous in t (and therefore in $z(t)$) is not a problem in our differential inclusion setting. It should be clear that in that case there is still a unique solution to the differential inclusion (13).

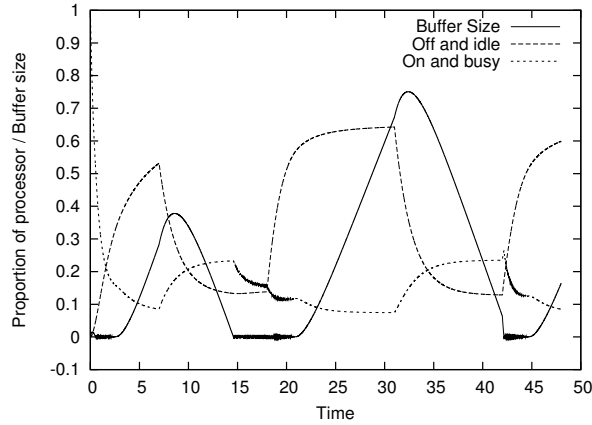


Figure 5: Limit dynamics for volunteer computing under day and night availability.

On Figure 5 we can observe two kinds of non-differentiable points. The first ones are the points representing the change from day time to night time. The other ones occur when the buffer becomes empty. The small oscillations of the buffer size around 0 and of the proportion of busy processors are just numerical integration artefacts (typical of numerical integration of differential inclusions). In both cases the exact solutions do not exhibit these jumps.

5.5 Join the Shortest Queue in Volunteer Computing

This example is similar to the one of Section 5.3 but with two identical time-homogeneous volunteer systems. Each time a packet arrives, it is routed to the system with the smallest number of packets. Here, the routing of packets introduces a new cause of non-smoothness: there is a threshold in the dynamics of the system when both backlogs are equal.

Figure 6 shows the behavior of the limit differential inclusion. Once again, the limit behavior is unique once the initial condition is given.

As expected, new non-differential points appear when both buffers are equal.

5.6 OSL dynamics: energy-aware distributed computing system

The previous examples have dynamics for which the drift function did not satisfy an OSL condition. However, OSL conditions are commonly assumed in the non-smooth system literature [5, 10]. In this section, we recall sufficient condition for a dynamical system to be OSL and provide some examples of dynamics that are OSL. Recall that a set-valued function F is OSL with constant $L \geq 0$ if for all $y, \bar{y} \in \mathbb{R}^d$ and $u \in F(y), \bar{u} \in F(\bar{y})$: $\langle y - \bar{y}, u - \bar{u} \rangle \leq L \|y - \bar{y}\|^2$.

Let us first examine the example of Section 5.3. Let $y = (b, a, u, C)$ and $\bar{y} = (\bar{b}, a, u, 0)$, then $\langle y - \bar{y}, f(y) - f(\bar{y}) \rangle = -\mu |b - \bar{b}|^2 + \gamma a(b - \bar{b}) - \gamma a$. When $b - \bar{b}$ is small enough and positive, this expression is of order $a(b - \bar{b})$ which is greater than $L \|y - \bar{y}\| = L(|b - \bar{b}|^2 + C^2)$. In fact, there is two types of non-smoothness in these equations. The first one is that the dynamics of C

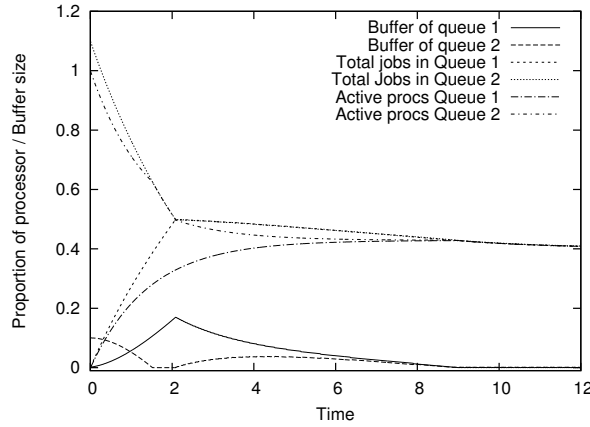


Figure 6: Join the shortest queue for volunteer computing

depends not continuously on the dynamics of C but in a OSL way (see below). The second type of discontinuity is that the dynamics of b depends non-smoothly on C . This latter discontinuity always leads to a term of order $(b - \bar{b})$ which is greater than $L \|b - \bar{b}\|^2$ whenever $b - \bar{b}$ is small enough. This is a general problem for the applicability of OSL: whenever the derivative of one coordinate depends non-smoothly on a other coordinate, the dynamics is never OSL. This is actually the case in all the examples of Sections 5.1 to 5.5.

5.6.1 Sufficient conditions to prove OSL

The next lemma links conditions on the drift function f to its set-valued counterpart.

Lemma 9. *Let $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ be a single valued function and F be the convex set-valued function associated to f , defined by*

$$F(y) = \bigcap_{\epsilon > 0} \overline{\text{conv}} \{f(z) : \|z - y\| \leq \epsilon\}.$$

where $\overline{\text{conv}}$ denotes the convex hull of a set. Then:

- (i) F is OSL with constant L iff f is OSL with constant L .
- (ii) If f is lipschitz of constant L , then F is OSL of constant L .

Sketch of proof. (i) – The proof of (i) is straightforward from the definition of OSL. The main idea is to use the fact that since \mathbb{R}^d is a d dimensional space and by the definition of F , any $u \in F(y)$ can be written as the limit of a convex combination of d points $u_1 \dots u_d$ such that $u_i = f(y_i)$ and $\|y - y_i\| \leq \epsilon$. Using a similar combination for $\bar{u} \in F(\bar{y})$ and playing with the coefficient of the convex combination leads to (i).

(ii) – The proof of (ii) comes from the Cauchy-Schwartz inequality. If f is Lipschitz with constant L , then for all $y, \bar{y} \in \mathbb{R}^d$: $\langle y - \bar{y}, f(y) - f(\bar{y}) \rangle \leq \|y - \bar{y}\| \|f(y) - f(\bar{y})\| \leq L \|y - \bar{y}\|^2$. This shows that f is OSL. By (i), this implies that F is OSL. \square

In order to show that a particular dynamic is OSL, we can also use the fact that the sum of two OSL functions is OSL. However, unlike in the case of Lipschitz functions, the product or the composition of two OSL functions is not necessarily OSL.

Lemma 10. *If F_1 and F_2 are two OSL with constant L_1 and L_2 , then $F_1 + F_2$ is OSL with constant $L_1 + L_2$, where $F_1 + F_2$ is defined by:*

$$(F_1 + F_2)(y) = \{u_1 + u_2 : u_1 \in F_1(y), u_2 \in F_2(y)\}.$$

The basic example of a OSL function is the fluid limit of a single M/M/1 queue. Let us assume that jobs arrive in the system at rate $N\lambda > 0$ and are served by a server with rate $N\mu > 0$. Let $Y^N(t)$ be the number of packet at time t in the queue rescaled by $\frac{1}{N}$ and let assume that $Y^N(0) = x$. The drift of the system is $f(y) = \lambda - \mu$ if $y > 0$ and $f(0) = \lambda$ (that can be extended to $f(y) = \lambda$ for $y < 0$). The set-valued drift corresponding to f is $F(y) = \{\lambda - \mu\}$ if $y > 0$, $F(y) = \{\lambda\}$ if $y < 0$ and $F(0) = [\lambda - \mu; \lambda]$. This function F satisfies a OSL condition with constant 0. Indeed, let $y, y' \in \mathbb{R}$. If y and y' are both positive or both negative, $f(y) - f(y') = 0$. If $y > 0$ and $y' \leq 0$,

$$\langle y - y', f(y) - f(y') \rangle = \langle y - y', -\mu \rangle \leq 0 \leq 0 \|y - y'\|^2. \quad (14)$$

which is less than 0 since $y > y'$. Therefore, the dynamics of a M/M/1 queue is OSL with constant 0 and one can apply Theorem 7 with $K_T = \lambda$, $I(N) = 1/N$, $J(N) = 0$ and any $L > 0$. Note that Theorem 7 requires $L > 0$ because of a term in e^{2LT}/\sqrt{L} that comes from $I(N) \sum_{i=1}^{T/I(N)} (1 + 4LI(N))^i \leq \exp 4LT/(4L)$. When $L = 0$, this term can be replaced by T . The dependence in T of the constants A_T , B_T and C_T is e^{2LT} when $L > 0$ and T when $L = 0$.

Let us now consider a more sophisticated computing model made of a single buffer and N computing resources. Jobs arrive at rate $N\lambda$ in the buffer (where λ may vary) and $C^N(t)$ is the number of jobs in the buffer rescaled by N . The buffer size is upper bounded by C_{\max} . In order to optimize energy consumption, the computing resource can compute at rate μ_1 or $\mu_2 > \mu_1$. If there is a proportion $M_1^N(t) = m_1$ resources at speed μ_1 and $M_2^N(t) = m_2 = 1 - m_1$ resources at speed μ_2 , the drift of $C^N(t)$ is $f_c(c, m_1, m_2) = \lambda - m_1\mu_1 - m_2\mu_2$ if $0 < c < C_{\max}$, $g(0, m_1, m_2) = \lambda$ and $f_c(C_{\max}, m_1, m_2) = -m_1\mu_1 - m_2\mu_2$. A computation similar to (14) shows that f_c is OSL.

The computing resources can not communicate and each resource updates its speed independently of the others as a function of the load $C^N(t)$ of the system. To do so, each resource measures the size of the queue $C^N(t)$ at rate γ . If the size of the buffer is c and the speed of the resource is μ_i , the speed is set to *fast* (*i.e.* μ_2) with probability $p_i(c)$ (and is set to *slow* with probability $1 - p_i(c)$). Therefore, the drift of $M_1^N(t)$ is $f_1(c, m_1, m_2) = \gamma(-m_1p_1(c) + m_2(1 - p_2(c)))$ and the drift of $M_2^N(t)$ is $f_2(c, m_1, m_2) = -f_1(c, m_1, m_2)$. If $b \mapsto p_i(c)$ is lipschitz, the functions f_1 and f_2 are both lipschitz. This shows that the drift of the whole system (which is the sum of f_c , f_1 and f_2) is a OSL function.

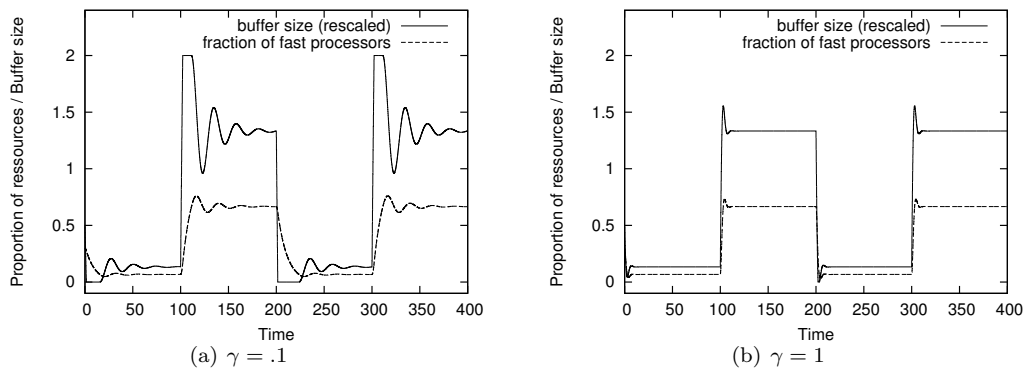


Figure 7: Differential inclusion dynamics of the energy-aware distributed systems for $\gamma = .1$ and $\gamma = 1$ and $C_{\max} = 2$. The y axis represents proportion of resources that are fast (*i.e.* $m_2(t)$) or the buffer size, depending on the curve.

In practical situation, measuring c and changing power of a computing resource is costly in terms of time and energy. Thus, the choice of γ is a compromise between the reactivity of the system and the time and energy spent to adapt the power to the load. Figure 7 shows two examples of the evolution of the differential inclusion dynamics for two values of γ : $\gamma = .1$ and $\gamma = 1$. The parameters of the system are $\mu_1 = .5$, $\mu_2 = 2$ and $C_{\max} = 2$. On each figure, $p_i(c) = c/C_{\max}$. When $t \in [0; 100]$ or $t \in [200; 300]$, we set $\lambda = .6$ while for $t \in [100; 200]$ or $t \in [300; 400]$, $\lambda = 1.5$.

Figure 7(a) shows that when γ is small, a brutal change of the arrival rate leads to a long period of oscillation of the system. When going from $\lambda = .6$ to 1.5, some jobs are lost at the beginning. Then, the system overreacts before stabilizing. When $\gamma = 1$ as in Figure 7(b), these oscillations have disappeared, except for a small burst at the beginning.

References

- [1] V. Acary and B. Brogliato. *Numerical methods for nonsmooth dynamical systems: applications in mechanics and electronics*. Springer Verlag, 2008.
- [2] F. Baccelli, D. McDonald, and J. Reynier. A mean field model for multiple tcp connections through a buffer implementing red. *Performance Evaluation*, 49, 2002.
- [3] M. Benaim, J. Hofbauer, and S. Sorin. Stochastic approximation and differential inclusions. *SIAM J. Control and Optimization*, 44(1):328–348, 2005.
- [4] Michel Benaim and Jean-Yves Le Boudec. A Class Of Mean Field Interaction Models for Computer and Communication Systems. *Performance Evaluation*, 65(11-12):823–838, 2008.
- [5] F. Bouchut and F. James. One-dimensional transport equations with discontinuous coefficients. *Nonlinear Analysis*, 32(7):891, 1998.
- [6] J.G. Dai. On positive Harris recurrence of multiclass queueing networks: A unified approach via fluid limit models. *The Annals of Applied Probability*, 5(1):49–77, 1995.
- [7] N. Gast, B. Gaujal, and J.-Y. Le Boudec. Mean field for Markov Decision Processes: from Discrete to Continuous Optimization. *INRIA RR*, 7239, 2010.
- [8] M. Kunze. Non-smooth dynamical systems. *Lecture notes in mathematics*, 2000.
- [9] T.G. Kurtz. *Approximation of population processes*. Society for Industrial Mathematics, 1981.
- [10] F. Lempio. Euler’s method revisited. *Proceedings of the Steklov Institute of Mathematics*, 211(4):429–449, 1995.
- [11] W. Rudin. *Functional Analysis*. McGraw-Hill Science, 1991.
- [12] W.J. Stewart. *Numerical solution of Markov chains*. CRC, 1991.

A Proofs of Theorem 5 and Theorem 7

The idea of the proofs is to write the evolution of the system as a stochastic approximation algorithm. The state of the system at time $k + 1$ can be written as

$$\begin{aligned} Y^N(k+1) &= Y^N(k) + I(N)f(Y^N(k)) + (f^N(Y^N(k)) - I(N)f(Y^N(k))) \\ &\quad + (Y^N(k+1) - Y^N(k) - f^N(Y^N(k))). \end{aligned} \quad (15)$$

where f is defined in (5). This equation can be seen as an Euler discretization of the DI (8) plus two error terms:

- $E^N(k) \stackrel{\text{def}}{=} f^N(Y^N(k))/I(N) - f(Y^N(k))$ which by Equation (5) is such that $\|E^N(k)\| \leq J(N)$;
- $U^N(k+1) \stackrel{\text{def}}{=} (Y^N(k+1) - Y^N(k) - f^N(Y^N(k)))/I(N)$ which by Equation (4) is such that $\mathbb{E}(U_{k+1}^N | \mathcal{F}_k^N) = 0$ where \mathcal{F}_k^N denotes the filtration associated with the process $(Y^N(k))_k$. Moreover, by Equation (7), $\mathbb{E}\left(\|U_{k+1}^N\|^2 | \mathcal{F}_k^N\right) \leq b$.

Using this notation, and the definition of $F(y)$ by Equation (10), Equation 15 can be rewritten

$$Y^N(k+1) \in Y^N(k) + I(N) (F(Y^N(k)) + E^N(k) + U^N(k+1)). \quad (16)$$

Equation (16) is called a *stochastic approximation with constant step size* associated to the DI (8). The term *constant step size* comes from the fact that $I(N)$ does not vary with time. Both proofs of Theorem 5 and Theorem 7 are based on the convergence of such stochastic approximation (16) as N goes to infinity.

A.1 Proof of Theorem 5

This proof is inspired by classical proofs on differential inclusions, for example the proof of Theorem 2.2.1 of [8]. It is somewhat similar to the proof of Theorem 4.2 of [3] although there are two main differences. First, we focus on *constant step size* and are interested in the convergence over a finite time-horizon. Second, we do not need any assumption on the boundedness of the stochastic process since this is a consequence of the hypothesis in the finite time-horizon.

The idea of the proof is to show that for all sub-sequence of \bar{Y}^N , there exists a sub-sequence $\bar{Y}^{\sigma(N)}$ (of this sub-sequence) such that $d(\bar{Y}^{\sigma(N)}, \mathcal{D}_T(y_0)) \xrightarrow{\text{a.s.}} 0$. In all that follows, let $\bar{Y}^{\sigma(N)}$ be a sub-sequence of \bar{Y}^N . In order to simplify the notations and because we will take several sub-sequences of sub-sequences, we omit the σ in the notation and we denote all sub-sequences \bar{Y}^N . In the first part of the proof, we consider the problem from a probabilistic point of view in order to make sure that the random part of the process goes almost surely to 0. Then we consider the problem from a trajectorial point of view using analytic arguments.

Developing the recurrence (16), the value of $Y^N(k+1)$ is equal to:

$$Y^N(k+1) = Y^N(0) + \sum_{i=0}^k I(N)f(Y^N(i)) + I(N) \sum_{i=0}^k E^N(i) + I(N) \sum_{i=0}^k U^N(i+1). \quad (17)$$

We define two functions $Z^N(t)$, and $V^N(t)$ to be piecewise affine functions such that for all $t = kI(N)$, $Z^N(t) = \sum_{i=0}^{k-1} I(N)f(Y^N(i))$ and $V^N(t) = \sum_{i=0}^{k-1} I(N)U^N(i+1)$. By Kolmogorov's inequality for martingales and since $\mathbb{E}(\|U^N(k+1)\|^2 | \mathcal{F}_k^N) \leq b$

$$\mathcal{P} \left(\sup_{0 \leq t \leq T} \|V^N(t)\| \geq \epsilon \right) = \mathcal{P} \left(\sup_{0 \leq k \leq \lceil \frac{T}{I(N)} \rceil} \left\| I(N) \sum_{i=0}^{k-1} U^N(i+1) \right\| \geq \epsilon \right) \leq \frac{I(N)T}{\epsilon^2} b. \quad (18)$$

This shows that $\sup_{0 \leq t \leq T} \|V^N(t)\|$ converges in probability to 0. Therefore, there exists a sub-sequence of V^N such that $\sup_{t \leq T} \|V^N(t)\|$ converges almost surely to 0.

We now reason from a trajectorial point of view. Let us now consider a trajectory $\omega \in \Omega$ of the system such that $\sup_{t \leq T} \|V^N(t)\|$ converges to 0. In particular, this implies that $\|V^N(t)\|$ is bounded for all N and t : $\sup_{N, 0 \leq t \leq T} \|V^N(t)\| \leq d < \infty$. Using (17) and since $\|F(y)\| \leq c(1 + \|x\|)$, for all $k \leq T/I(N)$, $\|Y^N(k+1)\|$ can be bounded by:

$$\begin{aligned} \|Y^N(k+1)\| &\leq \|Y^N(0)\| + \sum_{i=0}^k I(N)c(1 + \|Y^N(i)\|) + \sup_{N,t} \|V^N(t)\| \\ &\leq \|Y^N(0)\| + ckI(N) + d + \sum_{i=0}^k I(N) \|Y^N(i)\| \\ &\leq (\|Y^N(0)\| + cT + d) \exp(cT) / c, \end{aligned} \quad (19)$$

where we used the discrete Gronwall lemma and the fact that $kI(N) \leq T$.

Once we know that $\sup_{N, 0 \leq t \leq T} \|Y^N(t)\|$ is bounded, the rest of the proof can be adapted from classical results on the convergence of the Euler approximation for differential inclusions, see [8] for

example. There exists $e > 0$ such that $\sup_{N, 0 \leq t \leq T} \|Y^N(t)\| \leq e$. Thus $\|f(Y^N(k))\| < c(1+e) < \infty$. This shows that the functions Z^N are lipschitz with constant $c(1+e)$. Thus the sequence of functions $(Z^N)_N$ are equicontinuous and bounded. Therefore by the Arzela-Ascoli theorem, for all sub-sequence of $(Z^N)_N$, there exists a sub-sequence that converges to some $z : [0; T] \rightarrow \mathbb{R}^d$. In the following, we will show that z is a solution of (3) which shows that $d(Z^N, \mathcal{S}_T(y_0)) \rightarrow 0$. As $\|Z^N - Y^N\| = \|V^N\| \rightarrow 0$, this implies that $d(Y^N, \mathcal{S}_T(y_0)) \rightarrow 0$. To prove this, we will construct a function φ such that:

- (i) for all t : $z(t) = z(0) + \int_0^t \varphi(s) ds$;
- (ii) for almost every t : $\varphi(t) \in F(z(t))$.

Let $\varphi^N(t)$ be a step function, constant on the intervals $[kI(N), (k+1)I(N))$ and such that for $t = kI(N)$, $\varphi^N(t) = f(Y^N(k))$. Therefore, the sequence φ^N is bounded in $L_2([0; T], \mathbb{R}^d)$. Thus, there exists a sub-sequence of φ^N converging weakly in L_2 to a function φ . Since L_2 is a reflexive space, if a sequence of functions φ^N converges to φ , this means that for all function v , there exists a sub-sequence of φ^N such that $\langle v, \varphi^N \rangle \rightarrow \langle v, \varphi \rangle$. Let $\xi \in \mathbb{R}^d$ and $t \in [0; T]$. Let the function v be defined by $v(s) \stackrel{\text{def}}{=} \xi$ for $s < t$ and $v(s) \stackrel{\text{def}}{=} 0$ for $t \leq s$. Since φ^N converges weakly to φ and $Z^N(t) \rightarrow z(t)$, we have:

$$\begin{aligned} \langle Z^N(t), \xi \rangle &\rightarrow \langle z(t), \xi \rangle \\ \langle Z^N(t), \xi \rangle &= \langle Z^N(0), \xi \rangle + \left\langle \int_0^t \varphi^N(s) ds, \xi \right\rangle \\ &= \langle Z^N(0), \xi \rangle + \langle \varphi^N, v \rangle \\ &\rightarrow \langle z(0), \xi \rangle + \langle \varphi, v \rangle \\ &= \left\langle z(0) + \int_0^t \varphi(s) ds, \xi \right\rangle. \end{aligned}$$

As this is true for all $\xi \in \mathbb{R}^d$, this shows that z is absolutely continuous: $z(t) = \int_0^t \varphi(s) ds$.

It remains to show that for a.e. t , $\phi(t) \in F(z(t))$. Let t^N denotes the greater multiple of $I(N)$ less than t ($t^N \stackrel{\text{def}}{=} \lfloor t/I(N) \rfloor I(N)$). Using that $f(Y^N(k)) \leq c(1+e)$ and that z^N converges uniformly to y , for all $\delta > 0$, there exists N_0 such that $N \geq N_0$ implies $\|z(t) - Y^N(t^N)\| \leq \delta$.

Defining $F^\delta(y) \stackrel{\text{def}}{=} \cup_{z: \|z-y\| \leq \delta} F(z)$, this shows that $\phi^N(t) \in F^\delta(z(t))$. Since F is convex and z bounded, $\{\alpha \in L^2 : \alpha(t) \in F^\delta(z(t))\}$ is convex and closed. This shows that this set is weakly closed (see [11], Theorem 3.12). Therefore, for all t , $\phi(t) \in F^\delta(t)$. As this is true for all δ and F is USC, this shows that for a.e. t , $\phi(t) \in \cap_{\delta > 0} F^\delta(t) = F(z(t))$. Thus, z is a solution of the DI. \square

A.2 Proof of Theorem 7

This proof is based on two Lemmas. Lemma 12 shows the rate of convergence of an Euler scheme without error to the solution of a differential inclusion. Lemma 13) shows that the gap between the deterministic Euler scheme and the stochastic approximation can be bounded in probability with explicit bounds. The combination of the two directly implies Theorem 7.

We define by y_k^N the Euler scheme associated with the differential inclusion (3) with step $I(N)$ starting in $Y^N(0)$. That is:

$$\begin{aligned} y_k^N &= Y^N(0) \\ y_{k+1}^N &= y_k^N + I(N) (f(y_k^N) + E^N(k)) \end{aligned} \quad (20)$$

for some $f(y_k^N) \in F(y_k^N)$. Again, $y^N(t)$ denotes the piecewise interpolation of y_k^N where the time is scaled by a factor $I(N)$. More precisely, $y^N(t)$ is a piecewise affine function such that for $k \in \mathbb{N}$: $y^N(kI(N)) = y_k^N$. Finally, we denote by $y(t)$ a solution of the DI (3) (which is unique if the DI is OSL by Proposition 4).

We first start by a technical lemma that bounds the growth of the functions.

Lemma 11. *If F is USC and OSL with constant L and if there exists c such that for all x , $\|F(x)\| \leq c(1 + \|x\|)$, then: if y_k^N is the piecewise interpolation of the Euler scheme (20) with $\|E^N(k)\| \leq J(N)$ and $y(\cdot)$ the unique solution of the DI starting in y_0 , then there exist a constant $K_T^y, K_T^{y^N}$ and $K_T^{Y^N}$ such that $\sup_{0 \leq t \leq T} \|y(t)\| \leq K_T^y$, $\sup_{0 \leq t \leq T} \|y^N(t)\| \leq K_T^{y^N}$ a.s. and $\sup_{0 \leq t \leq T} \sqrt{\mathbb{E}(\|f(\bar{Y}^N(t))\|^2)} \leq K_T^{Y^N}$ where*

$$\begin{aligned} K_T^y &\stackrel{\text{def}}{=} (\|y(0)\| + c) \frac{e^{cT}}{c} \\ K_T^{y^N} &\stackrel{\text{def}}{=} (\|y_0^N\| + (c + J(N))T) \frac{e^{cT}}{c} \\ K_T^{Y^N} &\stackrel{\text{def}}{=} (\|Y_0^N\| + (c + J(N))T + \sqrt{bT}) \frac{e^{ct}}{c}. \end{aligned}$$

Proof. By the definition of y_k^N of Equation (20) and since $\|F(y)\| \leq c(1 + \|y\|)$, we have

$$\begin{aligned} \|y_{k+1}^N\| &= \left\| y_0^N + I(N) \sum_{i=0}^k (f(y_i^N) + E^N(i)) \right\| \\ &\leq \|y_0^N\| + I(N) \sum_{i=0}^k (c(1 + \|y_i^N\|) + J(N)) \\ &= \|y_0^N\| + (k+1)I(N)(c + J(N)) + I(N)c \sum_{i=0}^k \|y_i^N\| \end{aligned}$$

Therefore, for all $k \leq T/I(N)$, by the discrete Gronwall's lemma, $\|y_k^N\| \leq K_T^{y^N}$ with $K_T^{y^N} \stackrel{\text{def}}{=} (\|y_0^N\| + (c + J(N))T) e^{cT}/c$. Similarly, $\|y(t)\| \leq \|y(0)\| + \int_0^t c(1 + \|y(s)\|) ds \leq (\|y(0)\| + ct) e^{cT}/c$.

Similarly to the computation for y_k^N , and focusing on $Y^N(k)$ leads to:

$$\|Y^N(k+1)\| \leq \|Y^N(0)\| + (k+1)I(N)(c + J(N)) + I(N) \left\| \sum_{i=0}^k U^N(i+1) \right\| + I(N)c \sum_{i=0}^k \|y^N(i)\|$$

which shows that $\|Y^N(k)\| \leq (\|Y_0^N\| + (c + J(N))T + I(N) \left\| \sum_{i=0}^k U^N(i+1) \right\|) e^{cT}/c$. Since $\mathbb{E}(U^N(k+1) | \mathcal{F}_k^N) = 0$, $\mathbb{E} \left(I(N) \left\| \sum_{i=0}^k U^N(i+1) \right\|^2 \right) \leq I(N) \sum_{i=0}^k \mathbb{E} \left(\|U^N(i+1)\|^2 | \mathcal{F}_i^N \right) \leq bkI(N)$. Using the Jensen inequality on $x \mapsto (1 + \sqrt{x})^2$, this implies that for all $k \leq T/I(N)$

$$\mathbb{E} \left(\|Y^N(k)\|^2 \right) \leq (\|Y_0^N\| + (c + J(N))T + \sqrt{bT})^2 \frac{e^{2ct}}{c^2}.$$

□

Lemma 12. *If F is USC and OSL with constant L and if there exists c such that for all y , $\|F(y)\| \leq c(1 + \|y\|)$, then: if $y^N(\cdot)$ is the piecewise interpolation of the Euler scheme (20) with $\|E^N(k)\| \leq J(N)$ and $y(\cdot)$ the unique solution of the DI starting in y_0 , then*

$$\sup_{0 \leq t \leq T} \|y^N(t) - y(t)\| \leq e^{LT} \|y_0^N - y_0\| + \sqrt{I(N)} A_T + \sqrt{J(N)} B_T$$

where

$$\begin{aligned} A_T &\stackrel{\text{def}}{=} e^{2LT} \sqrt{\frac{I(N)}{6} c^2(1+K) + \frac{K}{2L} (c(1+K) + J(N))} \\ B_T &\stackrel{\text{def}}{=} \frac{\sqrt{K} e^{2LT}}{\sqrt{L}} \end{aligned}$$

and $K_T \stackrel{\text{def}}{=} \max\{K_T^y, K_T^{y^N}\}$ with K_T^y and $K_T^{y^N}$ defined as in Lemma 11.

If y_0^N is bounded, then for T fixed, A_T is bounded for all N . This shows that the convergence of $y^N(t)$ to $y(t)$ is of order $O(\sqrt{I(N)})$. In fact, the rate of convergence of the Euler's method can be improved if the solution of the differential inclusion satisfies some properties. For example, if the solution has a finite number of point at which it is not differentiable the rate of convergence is $O(I(N))$, see for example [10], Theorem 4.1.

Proof. Let us fix N and let us call $t_i = iI(N)$. Let $r(t) \stackrel{\text{def}}{=} \|y^N(t) - y(t)\|^2$. Since $y^N(\cdot)$ is piecewise affine and $y(\cdot)$ is absolutely continuous, r is absolutely continuous and differentiable for almost every t . Therefore, for a.e. t such that $t_i \leq t < t_{i+1}$, we have

$$\begin{aligned} \dot{r}(t) &= 2 \langle y^N(t) - y(t), f(y^N(t_i)) + E^N(i) - f(y(t)) \rangle \\ &= 2 \langle y^N(t_i) - y(t), f(y^N(t_i)) - f(y(t)) \rangle + 2 \langle y^N(t) - y^N(t_i), f(y^N(t_i)) - f(y(t)) \rangle \\ &\quad + 2 \langle y^N(t) - y(t), E^N(i) \rangle \\ &= 2 \langle y^N(t_i) - y(t), f(y^N(t_i)) - f(y(t)) \rangle \\ &\quad + 2 \langle (t-t_i) (f(y^N(t)) + E^N(t_i)), f(y^N(t_i)) - f(y(t)) \rangle + 2 \langle y^N(t) - y(t), E^N(i) \rangle \end{aligned} \quad (21)$$

Equation (21) can be bounded by the OSL condition and is less than $2L \|y^N(t_i) - y(t)\|^2 = 2L \|y^N(t_i) - y(t_i) + y(t_i) - y(t)\|^2 \leq 4L(r(t_i) + \|y(t_i) - y(t)\|^2)$. Moreover, since y is a solution of the DI (8), $\|y(t) - y(t_i)\| = \left\| \int_{t_i}^t f(y(s)) ds \right\| \leq ((t-t_i)c(1+K))^2$. By Cauchy-Schwarz inequality, the two terms of Equation (22) are bounded by $2(t-t_i)(c(1+K) + J(N))2K + 4KJ(N)$.

Since $r(t)$ is absolutely continuous, $r(t_{i+1}) = r(t_i) + \int_{t_i}^{t_{i+1}} \dot{r}(t) dt$. Thus

$$r(t_{i+1}) \leq r(t_i) (1 + 4LI(N)) + \frac{4L}{3} I(N)^3 c^2 (1+K)^2 + I(N)^2 (c(1+K) + J(N)) 2K + 4KJ(N) I(N).$$

Using that $(1+4LI(N))^i \leq e^{4Lt_i}$ and that $\sum_{j=1}^i (1+4LI(N))^j = ((1+4LI(N))^{i+1} - 1)/(4LI(N)) \leq e^{4Lt_i}/(4LI(N))$, we get by a direct induction that:

$$r(t_i) \leq e^{4Lt_i} r(0) + \frac{e^{4Lt_i}}{4L} \left(\frac{4LI(N)^2}{3} c^2 (1+K)^2 + 2KI(N)(c(1+K) + J(N)) + 4KJ(N) \right)$$

This quantity is maximized for $t_i = T$. Therefore, we have

$$\sup_{0 \leq t \leq T} r(t) \leq e^{4LT} r(0) + I(N) e^{4LT} \left(\frac{I(N)}{6} c^2 (1+K) + \frac{K}{2L} (c(1+K) + J(N)) \right) + J(N) \frac{Ke^{4LT}}{L}.$$

This gives the results using $\sqrt{a+b+c} \leq \sqrt{a} + \sqrt{b} + \sqrt{c}$. \square

The following lemma states the convergence in probability of $Y^N(k)$ to the Euler scheme $y^N(k)$ with explicit probabilistic bounds. Together with Lemma 12, it shows that Y^N converges to the solution of the solution of (8) with explicit bounds (i.e. Theorem 7).

Lemma 13. *If F is USC and OSL with constant L and if there exists c such that for all y , $\|F(y)\| \leq c(1 + \|y\|)$, then if $y^N(\cdot)$ is the affine interpolation of the Euler scheme (20) and $Y^N(\cdot)$ the affine interpolation of the stochastic approximation (16) with*

- $\|E^N(k)\| \leq J(N)$
- $\mathbb{E}(U_{k+1}^N | \mathcal{F}_k^N) = 0$ and $\mathbb{E}(\|U_{k+1}^N\|^2 | \mathcal{F}_k^N) \leq b$

then for all T and all $\epsilon > 0$,

$$\mathcal{P} \left(\sup_{0 \leq t \leq T} |\bar{Y}^N(t) - y^N(t)| \geq \epsilon \right) \leq \frac{I(N)}{\epsilon^2} \frac{e^{2LT}}{2L} \left(2c^2 \left((1 + K_T^{y^N})^2 + (1 + K_T^{Y^N})^2 \right) + b \right).$$

where $K_T^{y^N}$ and $K_T^{Y^N}$ are defined as in Lemma 11.

In particular, this shows that $\lim_{N \rightarrow \infty} \mathbb{P}(\sup_{0 \leq t \leq T} |\bar{Y}^N(t) - y^N(t)| \geq \epsilon)$ goes to 0 with rate $I(N)$.

Proof. Using the equality $\|x + y + z\|^2 = \|x\|^2 + 2(\langle x, y \rangle + \langle x, z \rangle + \langle y, z \rangle) + \|y\|^2 + \|z\|^2$ and the definitions of $Y^N(k+1)$ and y_{k+1}^N (Eq. (16) and (20)), $\|Y^N(k+1) - y_{k+1}^N\|^2$ can be rewritten

$$\begin{aligned} \|Y^N(k+1) - y_{k+1}^N\|^2 &= \|Y^N(k) - y_k^N + I(N)(f(Y^N(k)) - f(y_k^N) + U^N(k+1))\|^2 \\ &= \|Y^N(k) - y_k^N\|^2 \\ &\quad + 2I(N)\langle Y^N(k) - y_k^N, f(Y^N(k)) - f(y_k^N) \rangle \end{aligned} \quad (23)$$

$$\begin{aligned} &\quad + 2I(N)\langle Y^N(k) - y_k^N + f(Y^N(k)) - f(y_k^N), U^N(k+1) \rangle \\ &\quad + I(N)^2 \|f(Y^N(k)) - f(y_k^N)\|^2 + I(N)^2 \|U^N(k+1)\|^2 \\ &\leq (1 + 2LI(N)) \|Y^N(k) - y_k^N\|^2 \end{aligned} \quad (24)$$

$$+ I(N)^2 \left(\|f(Y^N(k)) - f(y_k^N)\|^2 + \|U^N(k+1)\|^2 \right) \quad (25)$$

$$+ 2I(N)\langle Y^N(k) - y_k^N + f(Y^N(k)) - f(y_k^N), U^N(k+1) \rangle. \quad (26)$$

where we use the OSL to go from (23) to (24).

Let us define a real valued sequence $W^N(k)$ by $W^N(0) = 0$ and for all $k \geq 0$:

$$\begin{aligned} W^N(k+1) &= (1 + 2LI(N))W^N(k) + I(N)^2 \left(\|f(Y^N(k)) - f(y_k^N)\|^2 + \|U^N(k+1)\|^2 \right) \\ &\quad + 2I(N)\langle Y^N(k) - y_k^N + f(Y^N(k)) - f(y_k^N), U^N(k+1) \rangle \end{aligned}$$

Because of Equation (24)-(25)-(26), for all $k \geq 0$, $0 \leq \|Y^N(k+1) - y_{k+1}^N\|^2 \leq W^N(k+1)$ (almost surely). Moreover, since $\mathbb{E}(U^N(k+1) | \mathcal{F}_k^N) = 0$, the expectation of (26) knowing \mathcal{F}_k^N is also 0. This shows that:

$$\begin{aligned} \mathbb{E}(W^N(k+1) | \mathcal{F}_k^N) &= (1 + 2LI(N))W^N(k) + I(N)^2 \left(\|f(Y^N(k)) - f(y_k^N)\|^2 \right. \\ &\quad \left. + \mathbb{E}(\|U^N(k+1)\|^2 | \mathcal{F}_k^N) \right) \\ &\geq W^N(k). \end{aligned} \quad (27)$$

This shows that $W^N(k)$ is a positive sub-martingale. Therefore, by Doob's inequality, for $\epsilon > 0$, $\mathcal{P}(\sup_{0 \leq i \leq k} W^N(i) \geq \epsilon^2) \leq \mathbb{E}(W^N(k)) / \epsilon^2$. Moreover, using Equation (27), the definition $W^N(0) = 0$, Lemma 11 and the hypothesis (7) on $\mathbb{E}(\|U(k+1)\|^2 | \mathcal{F}_k^N)$, $\mathbb{E}(W^N(k+1))$ can be bounded by:

$$\begin{aligned} \mathbb{E}(W^N(k+1)) &= \mathbb{E}(\mathbb{E}(W^N(k+1) | \mathcal{F}_k^N)) \\ &\leq (1 + 2LI(N))\mathbb{E}(W^N(k)) + I(N)^2 \left(\mathbb{E}(\|f(Y^N(k)) - f(y_k^N)\|^2) + b \right) \\ &\leq (1 + 2LI(N))^{k+1} W^N(0) \\ &\quad + I(N)^2 \sum_{i=0}^k (1 + 2LI(N))^i \left(2c^2 \left((1 + K_T^{y^N})^2 + (1 + K_T^{Y^N})^2 \right) + b \right) \\ &= I(N) \frac{(1 + 2LI(N))^{k+1} - 1}{2L} \left(2c^2 \left((1 + K_T^{y^N})^2 + (1 + K_T^{Y^N})^2 \right) + b \right). \end{aligned}$$

where $K_T^{y^N}$ and $K_T^{Y^N}$ are defined in Lemma 11. Note that we use Jensen's inequality on $x \mapsto (1 + \sqrt{x})^2$ to show that $\mathbb{E}(\|f(Y^N(k))\|^2) \leq c^2(1 + K_T^{Y^N})^2$.

Taking the value for $k = T/I(N)$, using the preceding remarks and the fact that for $a > 0$, $(1 + a)^i \leq e^{ai}$, we have

$$\begin{aligned} \mathcal{P} \left(\sup_{0 \leq t \leq T} \|\bar{Y}^N(t) - y^N(t)\| \geq \epsilon \right) &\leq \mathcal{P} \left(\sup_{0 \leq i \leq T/I(N)} W^N(i) \geq \epsilon^2 \right) \\ &\leq \frac{I(N)}{\epsilon^2} \frac{e^{2LT}}{2L} \left(2c^2 \left((1 + K_T^{y^N})^2 + (1 + K_T^{Y^N})^2 \right) + b \right). \end{aligned}$$

□



Centre de recherche INRIA Grenoble – Rhône-Alpes
655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
<http://www.inria.fr>
ISSN 0249-6399